Indian Health Service

National Data Warehouse

Documatron

Customer's Guide and Reference Manual
Documatron Metadata Process

January 2004

Department of Health and Human Services

Indian Health Service

Information Technology Support Center (ITSC)

Documatron

Customer's Guide and Reference Manual

Table of Contents

VERSION CONTROL

DATE	VERSION	NOTES
January 2004	1.0	Initial publish.

Documatron

Customer's Guide and Reference Manual

Overview

What is Documatron?

Documatron is a process, storage, and display methodology for enterprise metadata, or the 'data about the data'. It incorporates industry-standard taxonomy, yet allows for enterprise-specific data requirements in a one-source, one update environment.

"Data about Data?"

Metadata is data describing data, detailing its use throughout the enterprise. While this is obviously critical to the input and dissemination of captured data, metadata historically has been relegated to a secondary place in an enterprise. Commonly, MS Word or Excel mapping documents provide the only real 'link' to the data that is used to process enterprise applications. Since this metadata is maintained separately from the 'source' data, that data is outdated or discordant within a very short period.

Industries have begun to realize the importance of maintaining up-to-date information that not only allows retrospective and prospective review, but also maintains standards for enterprise-wide use, e.g., Data Warehouses. Understanding how the data is used is essential to providing management and utilization reports.

Application of metadata standards in the development of data elements assures a consistent, sharable set of data described and named in a predictable way. National and international efforts are underway to produce this metadata standardization.

The ISO/IEC JTC 1 Subcommittee 14, Data Element Principles, has developed a six-part standard addressing the standardization of metadata. Specification and Standardization of Data Elements (ISO 11179) describes rules, principles and guidelines for classifying, attributing, describing, naming, and registering data elements.

The popularity of the metadata registry concept has led to the formation of a user group, the Metadata Registry Implementers Coalition (MDRIC). Pressure from these and other potential users has had an impact on vendors. Both Ontek and Oracle have developmental efforts underway for commercial products based on the 11179 standard. Many implementations have been built independently, yet in accordance with the standard. Notable collaborations among subject area domains include the environmental area, health data, air traffic control, and statistical data.

Committees, such as the CWM (Common Warehouse Model), continue to meet and develop standards for establishing and communicating interoperable data.

The Dublin Core Metadata Initiative (DCMI), founded in Dublin, Ohio, in 1995, is an open forum that began development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices.

The Federal Geographic Data Committee (FGDC) manages digital geographic data primarily for use in Geographic Information Systems (GIS), image processing systems, and other modeling software. Data collections (spatial data sets) can be searched through a single interface based on their descriptions, or "metadata."

Metadata, therefore, provides the semantic layer between the decision support system and the business end users, significantly enhancing the value of the data warehouse / data marts.

For further information you may refer to these sites:

Introduction to Metadata
http://www.getty.edu/research/conducting_research/standards/intrometadata/

Common Warehouse Model (CWM)
http://www.omg.org/cwm/

Dublin Core Metadata Initiative (DCMI)
http://dublincore.org/

Federal Geographic Data Committee (FGDC)
http://www.fgdc.gov/

The National Institute of Standards and Technology
(http://www.nist.gov/welcome.html)

What is a Metadata Registry?

Metadata is synonymous with data; it can be stored and processed by a repository in much the same way data is stored and processed in a database. Mutually, both metadata and real data must be made available in a way easily accessible to users. This will allow the users the ability to search the metadata, thus being able to understand and locate the real data they need. This process will decrease the number of instances of redundant data gathering and processing.

A metadata registry is a means of managing shareable data – a comprehensive source of reference information about data. It does not contain the real data; rather it provides knowledge on the definition, origin, source, and location of the data. It supports the standard-setting process by recording and disseminating data standards, which in turn can facilitate data sharing among organizations and users.

A registry can include links to additional information about the metadata. When used in concurrence with an information database, the registry provides users the knowledge to better understand the gathered information and thus makes the information much more useful to the user.

There are packages commercially available, such as MetaPro, that allow users to create their own metadata registries. These tools are designed to foster exchange of metadata and encourage reuse of metadata attributes of data elements and value domains. Many of these tools are develop using database management systems such as Oracle and MS Access.

Why Documatron?

Documatron is specifically designed for the Indian Health Service's National Data Warehouse, making it as fundamental to the design of the warehouse as the storage of patient registration and encounter information. Unlike an off-the-shelf package, Documatron is not an "interface" for storing metadata, but rather a fully integrated module within the Data Warehouse itself.

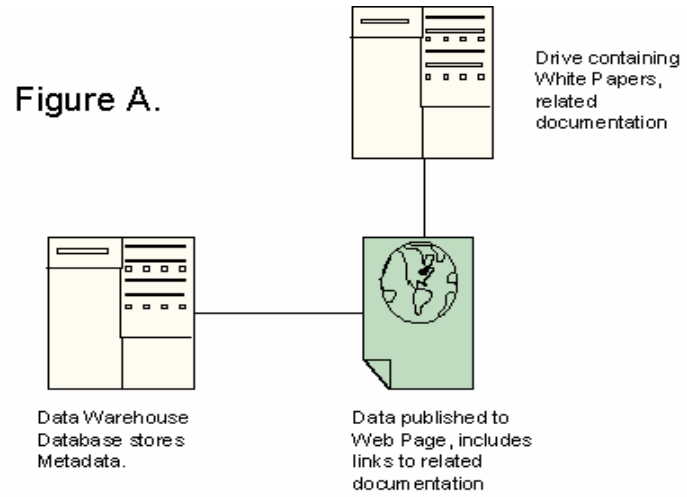
Unlike using a desktop application such as ACCESS or EXCEL, Documatron utilizes the built-in business and technical processes inherent within Database Management. And, since Documatron is part of the Data Warehouse, no additional system resources are needed to support the

product. The following items that were developed for the warehouse will include Documatron by default:

- System security protocols
- Database preventive maintenance and performance tuning
- System administration support
- Disaster Recovery

Figure **A**. is a simple representation of how Documatron is structured.

Figure A.



The IHS Documatron Process

Documatron Set-up for IHS

IHS' Data Warehouse will capture Registration and Encounter-based Patient data from sites that use both RPMS and non-RPMS systems, which is currently being loaded to NPIRS (National Patient Information Reporting System). IHS will use a generic interface with HL7 protocol standard for use by the IHS Data Warehouse for the purpose of exchanging data.

Each data element from all systems reporting data has been defined, along with any site-specific nuances, utilizing standard DB2 modeling terminology. These data element definitions form the metadata – the 'data about the data." In addition, specific system processes', e.g., ETL (Extract-Transform-Load) logic, and non-standard file processing (NFO) processes are also documented in these metadata tables.

The element definitions include:

- Name
- Description
- Lookup Tables
- White Paper Links
- ETL logic
- HL7 (mapping and definition)
- RPMS-specific processing rules, where applicable (for review and comparison)
- File Alignment (used for non-standard files)

The meta data tables and data will be generated (under DBA control) to the IHS Data Warehouse website for IHS DW users to view. In addition, different 'views' of the meta data will be available as pre-canned reports.

In addition to the data views and reports available on the IHS DW website, links to additional documentation, or "White Papers", will be maintained. Clicking on these links will direct users to additional details for those elements that require further explanation.